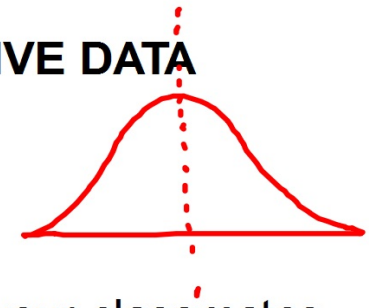
AP STAT: CHAPTER 4- QUANTITATIVE DATA



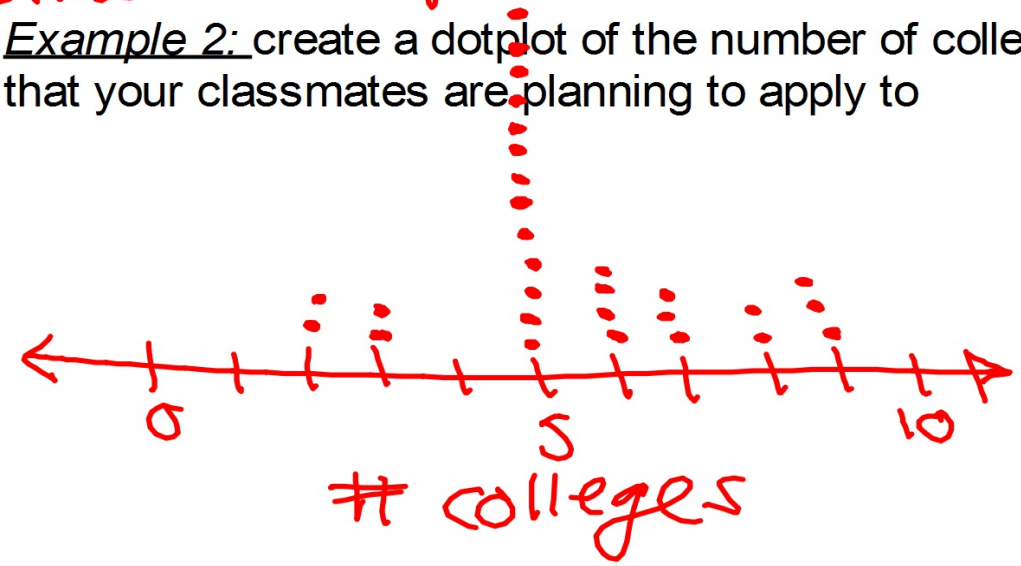
Quantitative Distributions:

1. Dotplots

Example 1: number of siblings of your classmates

* small samples

Example 2: create a dotplot of the number of colleges that your classmates are planning to apply to



1. Stemplot (aka Stem and Leaf Plot)

Example 2: Heights of classmates

60" = 5 ft.

Example 3: Babe Ruth's homerun totals each season for the Yankees:

54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 41, 34, 22

*small sets

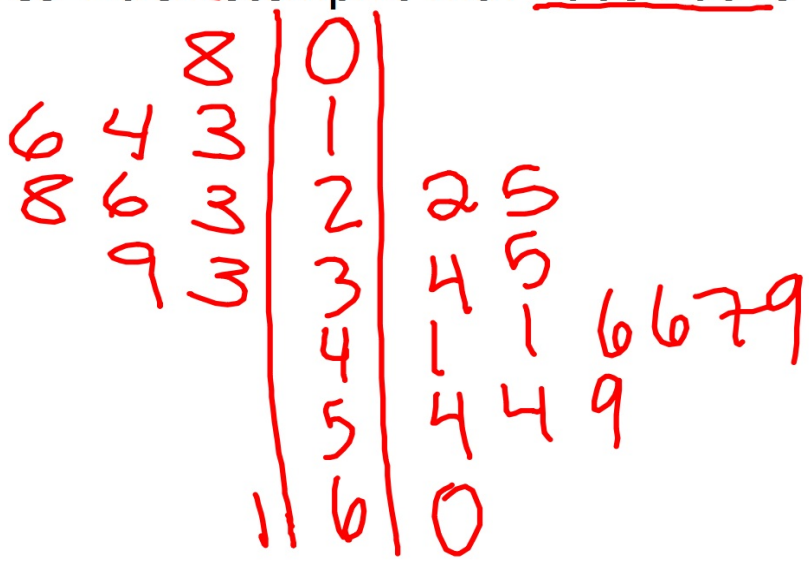
Create a stemplot

2	2 5
3	4 5
4	1 1 6 6 7 9
5	4 4 9
6	0

Example 3: Babe Ruth's homerun totals each season for the Yankees:
54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 41, 34, 22

Example 4: Roger Maris' homerun totals for the Yankees:
~~8, 13, 23, 33, 28, 16, 14, 39, 26~~, 61

Create a ~~**Back to Back**~~ Roger Maris stemplot with Babe Ruth's



Example 5: Age guesses

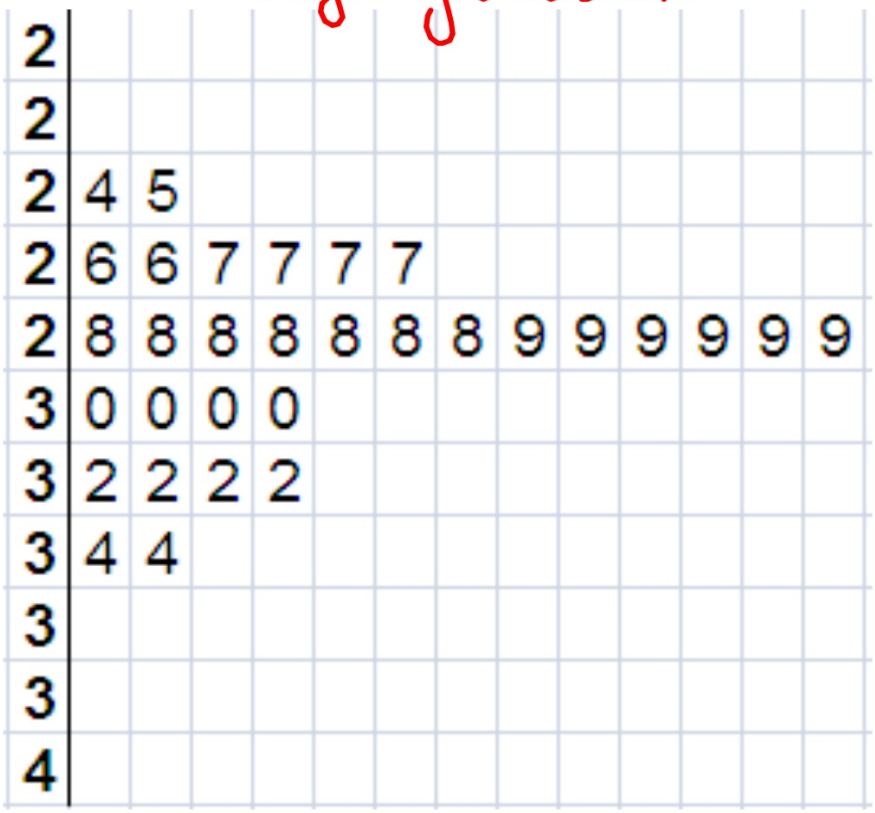
2	4 5 6 6 7 7 7 7 8 8 8 8 8 …
3	0 0 0 0 2 2 2 2 4 4
4	

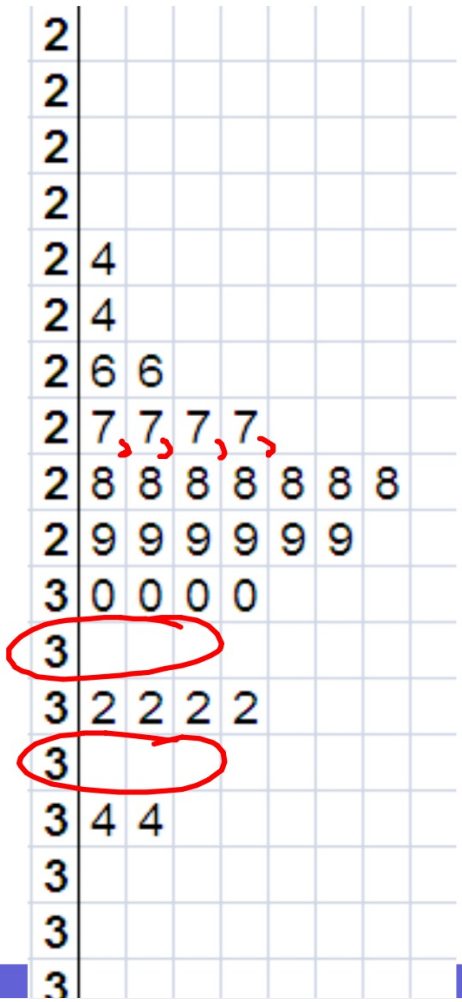
*clustered

Splitting stems

20–24	2 4
25–29	2 5 6 6 7 7 7 7 8 8 8 8 8 …
	3 0 0 0 0 2 2 2 2 4 4
	3
	4
	4

Age guesses




2
2
2
2
2 4
2 4
2 6 6
2 7 7 7 7
2 8 8 8 8 8 8 8
2 9 9 9 9 9 9
3 0 0 0 0
3
3 2 2 2 2
3
3 4 4
3
3
3

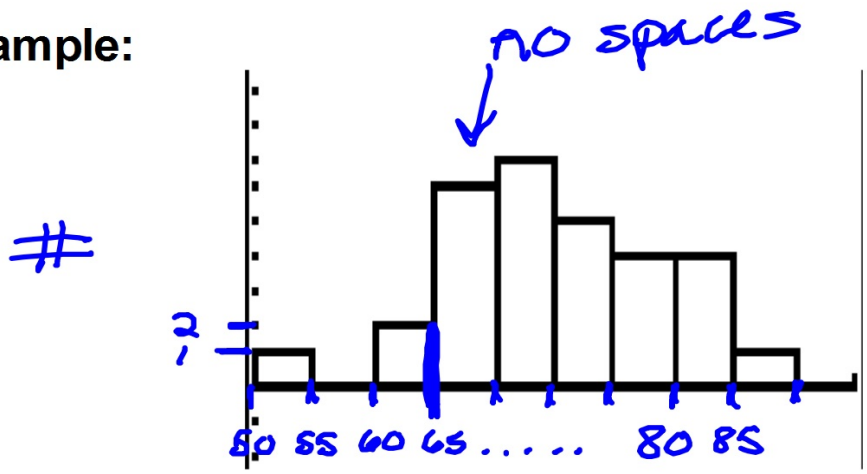
Complete the worksheet about Inputting, Transferring, and working with lists with a partner

When done, come up to the front desk and transfer the following lists to your calculator:

GPA
INCOM
SATMF
SATMM
TEST

Another quantitative distribution: HISTOGRAMS!

Example:



TEST SCORES

65

large samples

Histograms

Example: The following are a list of test scores on an exam. Let's create a histogram of these scores

40	61	66	68	74	77	84	91
45	62	66	69	75	78	84	95
49	64	66	69	76	80	85	96
51	64	67	70	76	81	85	96
53	64	67	71	76	81	86	99
57	64	68	72	76	82	87	
59	65	68	72	76	82	90	

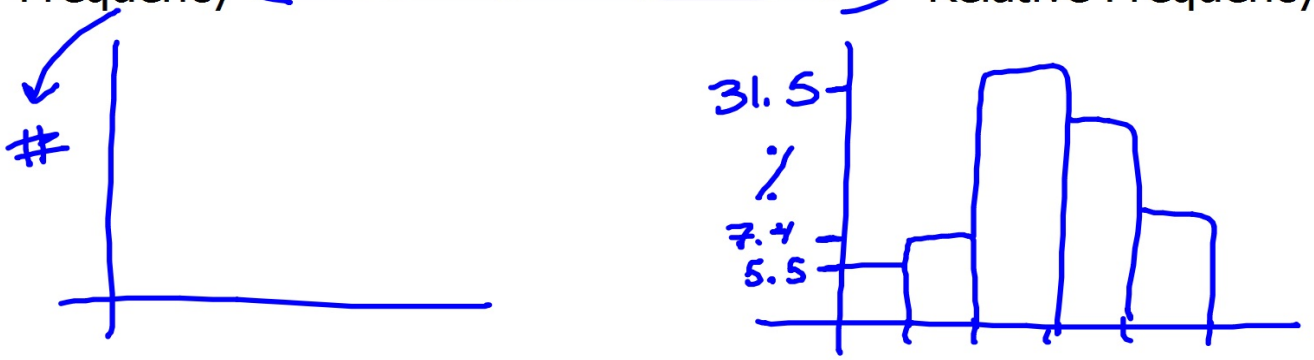
BINS =

Bar

%

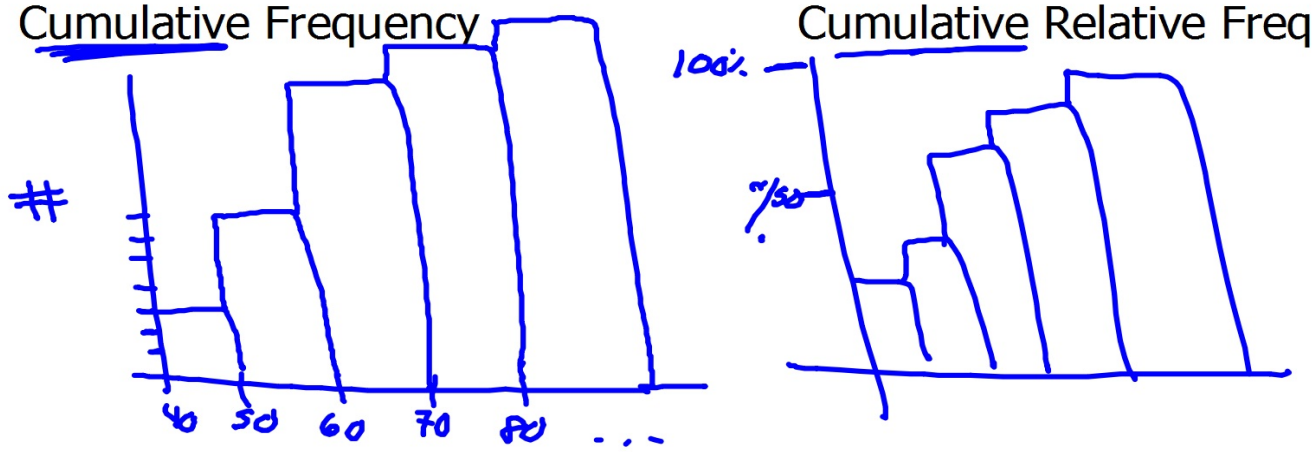
4 types of histograms:

Frequency ← same pic. → Relative Frequency



Cumulative Frequency

Cumulative Relative Freq



Using a chart is very helpful when creating the other types of histograms:

Bin	Frequency	Relative Frequency	Cumulative Frequency *	Cumulative Relative Freq. *
40-50	3		3	
50-60	4		7	
			.	
			⋮	
Total:				

Histograms on the calculator: see page 46 in the book for help

Example using the list TEST from page 1

Histogram Examples:

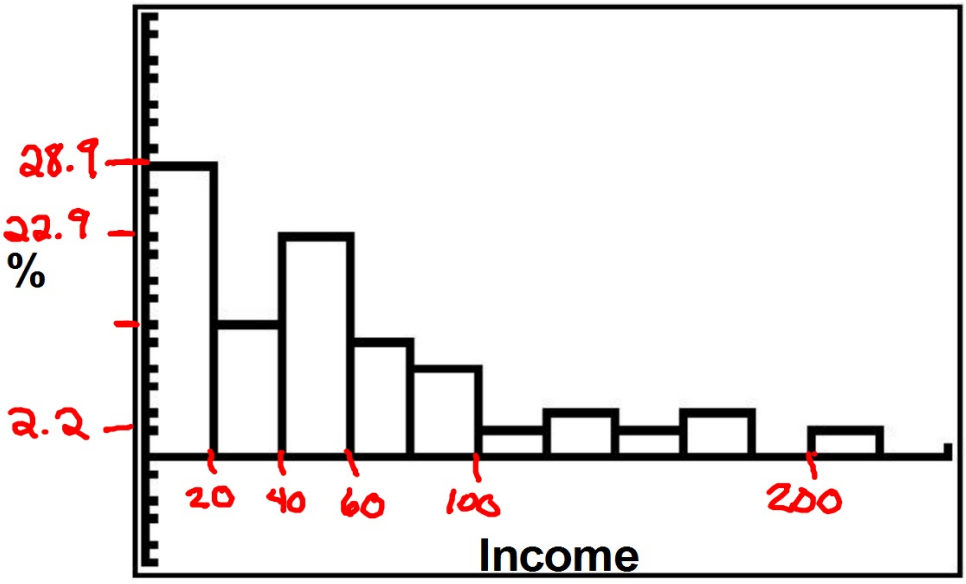
Example 1: Using the class data of AGES IN MONTHS (question 18), create a **frequency** histogram on your paper

Examples 2: Using the list INCOM, create a **relative frequency** histogram on your paper

Example 3: Using the list GPA, create a **cumulative frequency** histogram on your paper

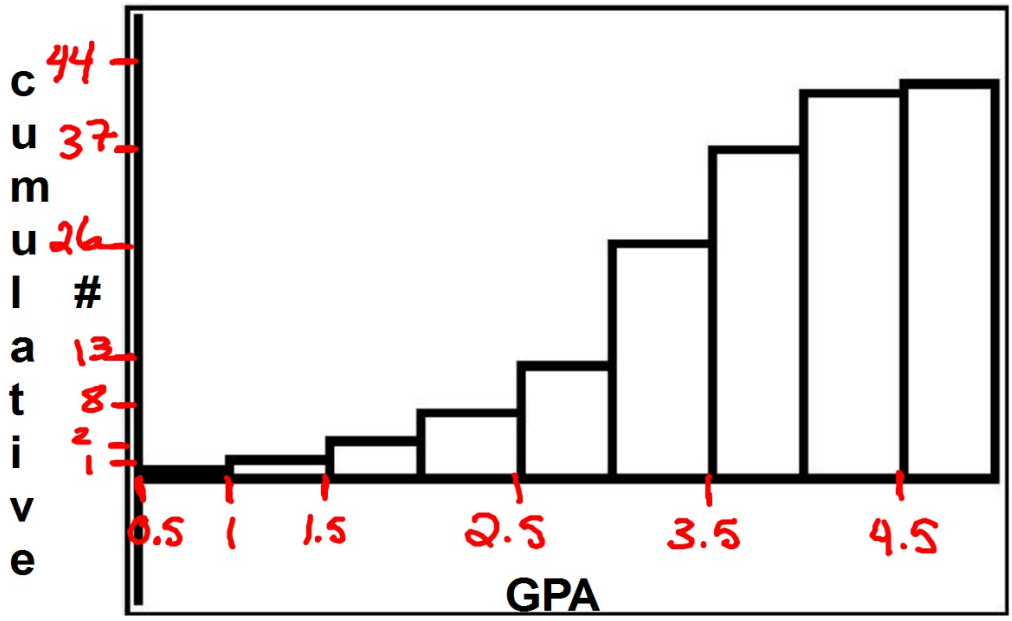
Answers to examples:

Example 2:



cumul. freq.

Example 3:



Transfer the list STATE to your calculators, then.....

Example 1: Using the class data of AGES IN MONTHS (question 18), create a **frequency** histogram on your paper

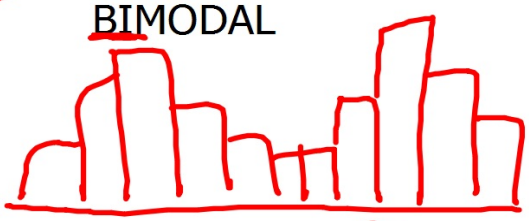
DESCRIBING DISTRIBUTIONS: Shape, Center, Spread

SHAPE

MODE: UNIMODAL

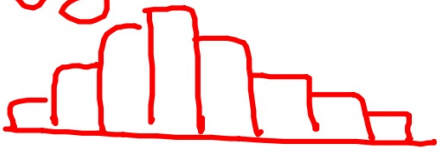
multimodal

BIMODAL



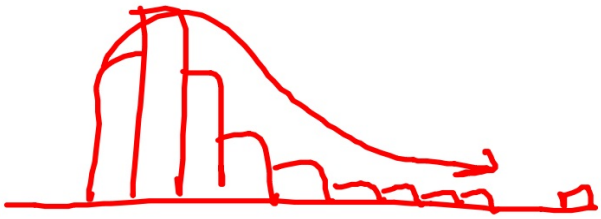
SHAPES: UNIFORM

roughly SYMMETRIC



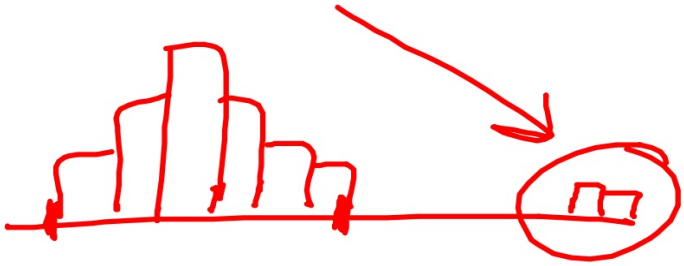
LEFT SKEWED

RIGHT SKEWED

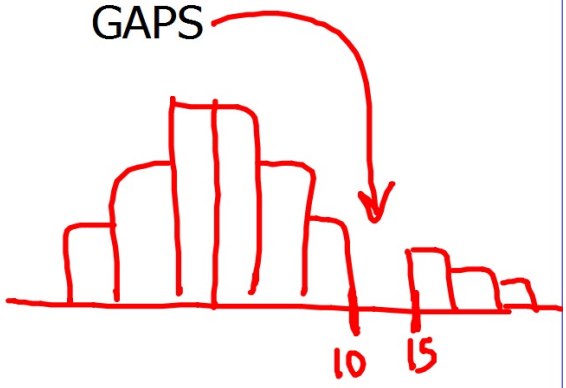


OTHER: OUTLIERS

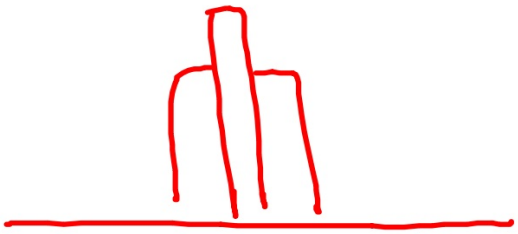
unusual observations



GAPS



CLUSTERED



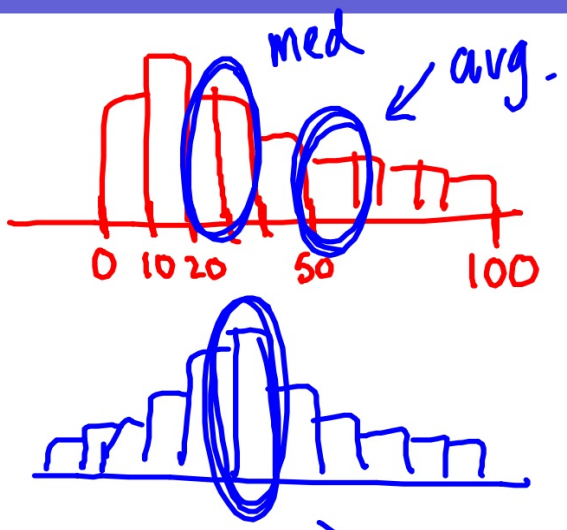
GRANULARITY

Ex: MC, IO?

consistent & understandable gaps

CENTER:

- approximate center
- mean or median



SPREAD:

RANGE: (lowest, highest)

Try the worksheet...

① unimodal, symmetric, range of (1, 12) and center @ ~7.

② unimodal, left skewed, range (50, 190), gap btw. 80 + 100, center ~ 150

③ uniform, range (3, 11) center ~ 7

Describing Distributions... numerically!

CENTER: Median

- middle obsevation
- Arrange data in order from least to greatest
- If an odd number of obs. find the middle number
- If an even number of obs. find the average of the middle two numbers

Median number of states visited (list STATE)

Order:

2 2 3 4 5 5 6 6 7 7 7 7 7 8 9 9 9 10 11 12 14 16 19 27 35

Med = Q2

Quartiles:

- medians of the lower and upper half of the data

Quartiles = Q1 = med. lower ½ = 5.5

Q3 = med. upper ½ = 11.5

5 Number Summary

A numeric description of the distribution:

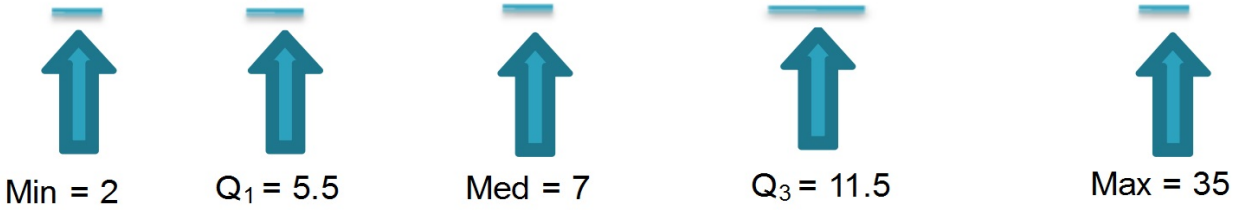
* Minimum
* 1st Quartile (Q_1)
* Median
* 3rd Quartile (Q_3)
* Maximum

middle 50%

5 Number Summary

- Put data in order

2 2 3 4 5 5 6 6 7 7 7 7 8 9 9 9 10 11 12 14 16 19 27 35



* on the calculator... 1-var stats

Measures of Spread

When we have chosen the median....

* Range = (Max, Min)

* Interquartile Range

$$IQR = Q_3 - Q_1$$

CENTER: ***Mean***

* Arithmetic average
* Add all observations together and divide by the number of observations.
* Formula:

* Read as “x bar”
* ***Must be written with the bar!!!!***
* on calculator: 1-var stats.... x-bar

Measure of Spread

When we have chosen Mean...

* *Standard Deviation* is a number that describes how much data vary or spread out.
* Uses difference of each data value from the mean.
* The higher the number the more the data is spread out.

Examples:

Finding Standard Deviation

1. Find the mean, $\bar{x}$

2. Find the difference of each data value from the mean $x_i - \bar{x}$

3. Square each difference $(x_i - \bar{x})^2$

4. Sum all the squared values

$$\sum_{i=1}^{n}(x_i - \bar{x})^2$$

5. Divide the sum bv the number of data values minus one, (n – 1)

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

This value is called the variance, s^2

6. Finally take the square root to find the standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Example

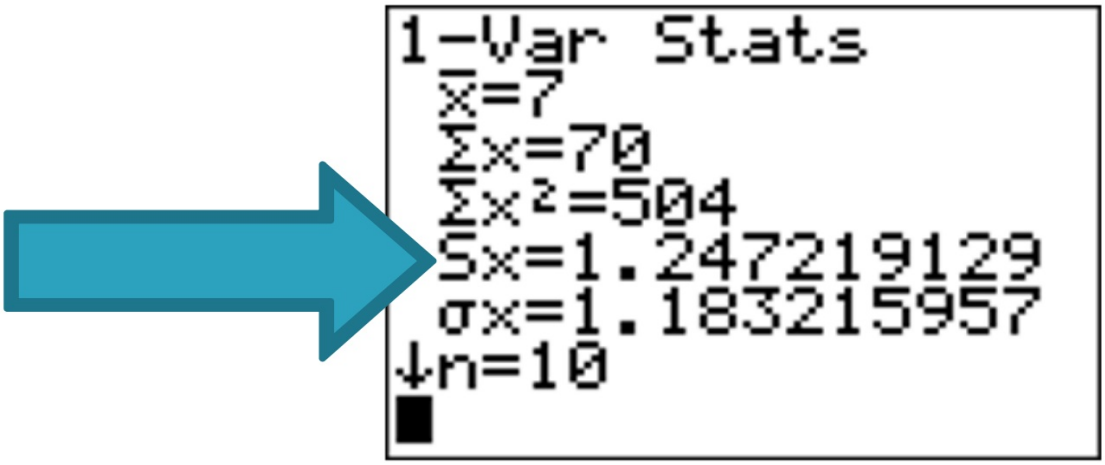
{6, 8, 9, 8, 7, 5, 6, 7, 6, 8}

$$\bar{x} = \frac{6+8+9+8+7+5+6+7+6+8}{10} = 7$$

y	6	8	9	8	7	5	6	7	6	8
$x-\bar{x}$										
$(x-\bar{x})^2$										

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 =$$

Or just do it on your calculator!!!



Summary:

CENTER:	MEDIAN	MEAN
SPREAD:	RANGE IQR 5# summary	STD. DEVIATION

Mean vs Median (in distributions)

* Symmetric:

mean $\approx$ median

* Left skewed:

mean $<$ median

* Right skewed:

mean $>$ median

* The mean is affected by outliers and the median is ***resistant*** to outliers

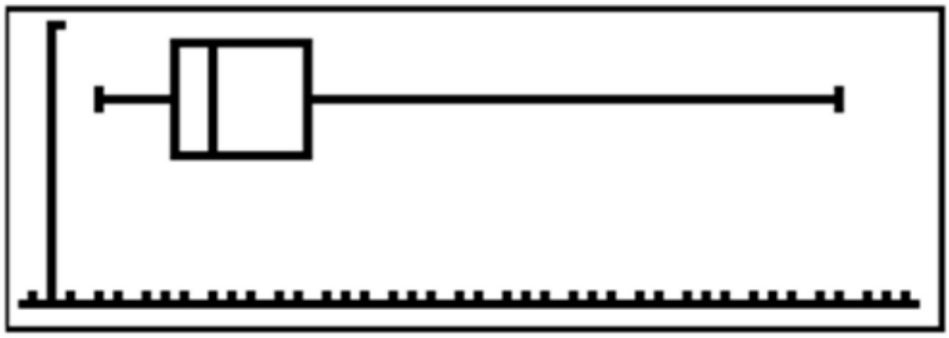
Ex. Ages in months

** ***Symmetric data => Mean and Std. Deviation***

** ***Skewed data => Median and IQR***

Boxplot

- Graphical display of 5 number summary



1.5 Rule for Outliers

- Any observation below the Lower Fence (LF) or above the Upper Fence (UF) is considered an outlier
- $LF = Q_1 - 1.5(IQR)$
- $UF = Q_3 + 1.5(IQR)$

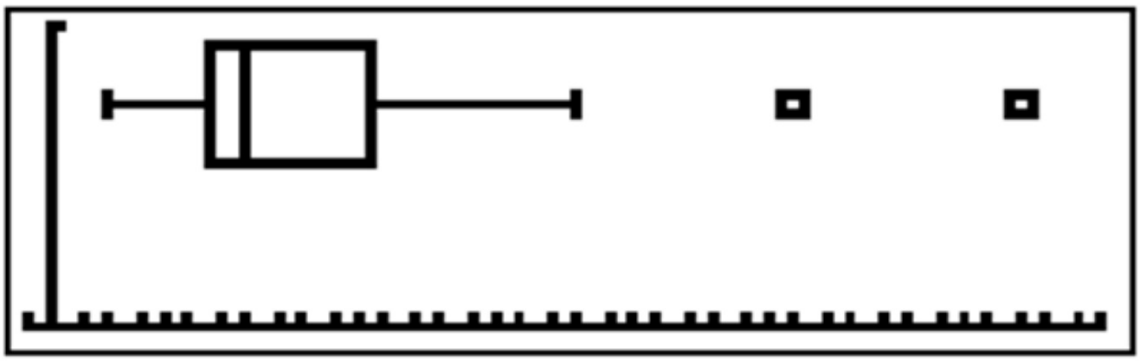
Example:

Using the AGES in months data, test to see if my age is an outlier

n = 29
min = 204
Q1 = 204.5
Med = 210
Q3 = 213
Max = 340

Modified Boxplot

- If outliers are present the whisker extends to the last value not an outlier
- The outlier is marked with an x, box, or point



Parallel boxplots:

- 2 or more boxplots, on same scale
- used to compare two or more similar variables
- on calculator:

Practice...